DOCUMENT RESUME

ED 275 173 FL 016 077

AUTHOR Shohamy, Elana
TITLE Input and Output in Language Testing.
PUB DATE 84
NOTE 10p.; In: Practice and Problems in Language Testing. Papers from the International Symposium on Language Testing (7th, Colchester, England, 1984); see FL 016 066.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Language Proficiency; *Language Tests; *Linguistic Theory; *Measurement Techniques; Second Language Learning; *Test Construction; *Testing Problems

ABSTRACT

This study criticizes current approaches to language test development for their restrictive views of how language should be measured. Language tests are based either on an input component (what goes into a test) or an output component (how should the test be measured). An integration of both components on language tests is advocated. The descriptions provided by Stern and Cummins are recommended for addressing the input issues, and measurement theories are suggested as sources for output. However, it is noted that because theories of language proficiency are not complete, many of the available sources may have only preliminary information useful for test construction, and that measurement theories still do not provide sufficient and full direction regarding the psychometric procedures which should be used to deal with communicative and authentic language. As a result, it is suggested that special efforts be made to advance these two components toward their integration in language testing. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275173

INPUT AND OUTPUT IN LANGUAGE TESTING

ELANA SHOHAMY
(TEL AVIV UNIVERSITY)

FL 016 077

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. Cuthane

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

INPUT AND OUTPUT IN LANGUAGE TESTING

Elana Shohamy (Tel Aviv University)

1. Introduction

The issue of quality tests has been discussed talked about, and written about for some time. Stevenson (1981) provides a distinction between "good" and "bad" language tests by pointing to two views of language testing - a "restricted" view which focuses on the technical aspects of tests, and an "expanded" view which focuses on the scientific aspect of such tests. He elaborates on those two views by stating that the restricted view refers to situations where the test constructors have little or not training and expertise in testing and measurement theories. They are only marginally concerned with aspects of validity, reliability and language testing theories. Furthermore, testing is perceived as a technical skill and is concerned mainly with the "how to do it". It focuses only on the visible content, construction and scoring of tests.

The "expanded" view, on the other hand, refers to a situation where there is awareness of a scientific approach to the study of language behaviour and measurement of language. Those who have an expanded view to testing are interested in various testing approaches such as essays, sociolinguistic elicitation procedures and their effects. Moreover, they are aware of the controversies regarding the nature of language and what is actually being measured when language ability is tested. The expanded view is therefore concerned with basic questions of validity and validation. It is less assured, less prescriptive and much more aware of the great complexity of what it is attempting to measure. It is based on a theory of what it means to know a language. Because of the incompleteness of such a theory it advocates the use of multiple procedures for language testing. Its prime concern is construct validity, while reliability is taken for granted. It asks itself "why test" and "what" exactly can and should be tested. The expanded view then calls for the development of testing instruments which are based on a clear definition of the domain of second language learning and what it means to know a language.

Stevenson (1982) claims that the restricted view is narrow since the emphasis is on the "how" of testing, that is, the techniques found in text books for constructing tests which are much less important than an understanding of the why and the what. Unfortunately, he claims, it is frequently the how that receives most of the attention from language teachers. Yet, any language test is necessarily dependent upon the purpose of the testing procedure, and upon how one answers the question of what it means to know a language.

This paper will provide a different distinction between two approaches to language testing where each approach is based upon a

different assumption - the synthesis of both approaches can result in quality language tests. It is claimed that in the field of language testing today, each of these approaches is used independently while the production of good tests requires the integration of both.

2. The Problem

In the past few years there have been serious discussions in the field of language testing about measurement issues such as test reliability, and validity, and little mention, if at all, on issues related to language learning and acquisition as they relate to language testing. Measurement experts are used to dealing with psychometric issues without finding a need to delve into the theory of the domain being tested. After all psychometrics is a well established field which devotes itself to such issues and there is an understanding that measurement experts do not need to know much about the domain being tested.

However, foreign language learning is different than other subjects. It is unique in several aspects: first, unlike other subjects it can be learned without formal teaching, thus in constructing tests there is not always a need to relate to and depend on a specific curriculum for deciding what to test. Second, unlike other school subjects, the process of second language learning as well as language proficiency have been researched and there are some findings in areas such as the successful language learner, attainable levels of proficiency, effects of social and individual factors on learning a foreign language, all of which have an impact on the testing of foreign languages. Thus, the language tester needs to know about the field of Applied Linguistics, the definitions of language and its knowledge, the importance of knowing how to use and communicate in a language, and that knowing a language is perceived as something more than discrete elements of lexicon and grammar.

Thus, even a well-trained measurement expert who is knowledgeable in modern and classical measurement theories in areas such as item analysis, reliability, validity and latent trait models will not be able to construct good language tests without knowing what it means to know a language. He needs to know what the various phases of language learning are and what the expected levels of foreign language proficiency at different stages are. Now knowing about these aspects makes it impossible for the measurement expert to construct tests which will be construct valid since he cannot know what should be tested, and more important, what the criteria for accuracy and correctness are. He may be testing the "wrong" things and may be scoring the tests based on the "wrong" criteria. One example to illustrate this phenomenon is the testing of pronunciation, or accent in foreign languages. For many years accent was tested in various language tests in a variety of ways. However, research of the past few years in second language acquisition has shown that only a very small portion of foreign language learners can fully acquire the foreign language accent and that there are various degrees of

accent proficiency that are attainable by the foreign language learner, depending on specific situations. Such information is essential for the tester to have in order to construct valid speaking tests. Still, this type of information by itself is clearly necessary but not sufficient for producing good tests of speaking. There is also a need to impose rigorous psychometric criteria on that information to convert it to a test. The test needs to differentiate and discriminate among a wide variety of speakers, all the items must relate to the construct being tested, it must give information on the likelihood of obtaining similar results upon a second administration of the same test, and it must be difficult enough for the good students and easy enough for low level students. It is clear from this example that the construction of quality language tests requires the integration of knowledge from these two domains - language learning and measurement theories.

Another example to illustrate the same point is the testing of sociolinguistic proficiency. Till today, sociolinguistic proficiency has been rarely tested. Most language tests still focus on linguistic aspects as the major criteria for test construction. However, the past few years have brought about an awareness of the use of language within a social context. It is apparent today that knowing a language means also knowing when, where, how and to whom to use it. Overlooking the state of the art in what knowing a language means today, implies the construction of tests which fail to tap the construct of language proficiency in its full and complete definition. For the person who is aware of sociolinguistic definitions, the construction of such tests will obviously involve imposing rigorous measurement criteria to convert this sociolinguistic information into tests of sociolinguistic proficiency.

Unfortunately, many of the tests used for arriving at conclusions about individuals are based on instruments which are incomplete either because they are not based on a fuller theory of knowing a language or they do not possess the necessary measurement properties. Thus, tests which are based on language learning theories without good psychometric theories are not sufficient in the same way that tests which possess measurement properties but are not based on complete language learning theories are insufficient.

The claim that is made here is that there is a need for greater reliance on and application of existing theories of language learning in order to define more fully the domain being tested. Any other way is restricted and narrow in its scope and inevitably leads to tests which lack construct validity. At the same time there is a need to impose rigorous measurement criteria on tasks which draw on language learning theories. Without these criteria this knowledge cannot serve as a valid basis for testing instruments. Thus knowledge of theories of second language learning on one hand and measurement theories on the other, are necessary but in themselves insufficient to produce quality language tests - there is a need for integration of the two domains.

3. Input and Output

The two domains - language learning and measurement have different, yet complementary roles and purposes in producing language tests. Language learning theories can bring the INPUT component into language testing i.e. defining what should be tested and what the appropriate language behaviours are. Measurement theories on the other hand can bring the OUTPUT component of the test i.e. how can language performance and proficiency turn into tests of these constructs.

3.1 The input component

Input in second language is concerned with the definition of language and what it means to know a language for given purposes. Second language test constructors must define proficiency levels as objectives or as standards by which to assess foreign language knowledge. The objectives or standards can serve as criteria for judgement.

Cummins and Stern (1981) provide a comprehensive view of some approaches to determine what it means to know a language. They distinguish among four distinct approaches to the phenomenon and claim that none of the four is really sufficient since there is a need of all the four to be combined for a fuller picture of language proficiency. The four approaches are: a) theoretical concepts; b) rating scale; c) formal proficiency tests; and d) inter-language research.

The four approaches will be briefly reviewed here with the understanding that each, and perhaps all four combined, could serve as a basis for defining the domain that can be used as the input component of a test.

The first approach is a theoretically based concept of proficiency. Three concepts are described here: a) proficiency defined in linguistic content. Up till the 1970's the theoretical linguistic model included phonology, vocabulary and grammar. More recently features of semantics, discourse and sociolinguistics have been added to the theoretical concept of proficiency. Today the concept of proficiency is expressed in communicative and not only in grammatical terms, but obviously do not exclude grammar; b) the second concept is the psychological or behavioural one. Such description of proficiency ranges from the abstract concept of "competence" to the more concrete term of listening, speaking, reading, and writing skills. At the most concrete level, proficiency is described as categories of language items, situations, speech functions and topics. Others define proficiency in terms of syllabus of items for specific groups of learners. c) A third concept that is made up of both linguistic and behavioural features is that of stages or levels of L-2 proficiency, which describes schemes involving different levels, each described in terms of the four language skills.

The second approach to proficiency is through describing

different proficiency levels by means of a rating scale. This approach is made on the basis of learner's practical knowledge at different stages. The FSI and Defense Language Institute have designed rating scales for the four language skills, which are divided in terms of communicative functional skills. Such rating scales describe expected standards for achievement levels in terms of the objectives to be aimed at. Stern and Cummins claim that rating scales can provide useful impressionistic descriptions of the stages and development of proficiency at different levels.

The third approach defines proficiency as a result of formal tests. Language proficiency tests imply a conception of proficiency that is based on empirical data. However, since researchers still have no definite answer to the question of the nature of language proficiency, some language tests results can do that. Unfortunately, they claim, most present language tests primarily measure the academically related aspect of proficiency and not the communicative and interpersonal components.

The fourth approach to proficiency is provided by the inter-language studies where the researcher examines closely what learners actually do in their L-2 performance. They refer to Selinker's work which describes the proficiency of L-2 learners as a linguistic system in its own right between the L-1 and the target language. The contribution of interlanguage research to this description of proficiency may lay in the concrete data that such studies provide on the development of proficiency. Such data can be extremely useful for providing a basis upon which one can plan and build tests. The work of Blum-Kulka and Olshtain (1984) in which they describe inter-language of speech acts of apologies and requests can also be a promising description of sociolinguistic data that can be extremely useful in constructing tests based on sociolinguistic input.

It is clear from the review above that there are various sources that language testers can draw upon in order to obtain the necessary input data. By focusing on the input aspects, the language tester can be confident that he is building on a solid foundation.

3.2 The output component

Input data is the foundation of a test but in itself is not sufficient for producing good tests. There is a need to convert the information described in the input into a test by providing empirical measurement data. The input data can serve as a basis for the language test which also needs to possess the appropriate measurement properties. These properties relate to the accuracy of the tasks and their ability to provide good judgement. The kind of measurement criteria imposed depend on the purpose of the tests and thus will differ from one test to another. However, all the criteria are intended to indicate that the testing instrument is capable of judging and making decisions about individuals' language knowledge. The main properties generally used for the output are the various types of reliability and validity as well as item analysis. It is suggested that the measurement criteria be imposed after the input has been defined.

3.3 The integration

The language tester's role is to integrate the input and the output components. The language tester needs to construct testing tasks based on the input, that lend themselves to testing and that will also possess the needed measurement properties.

4. Real-life examples

The type of work done in language testing in the past few years and some of the papers presented at this meeting, can serve as live examples to illustrate the points made above about input and output in language testing. Two types of tests seem to emerge in the field - input type tests and output tests. The input type tests are concerned primarily with the need to base tests on theories of language proficiency and with defining language as communicative interactions. The example to illustrate this is the entrance exam to British Universities (the TEAP test) as it was described in this meeting by Weir and Emmett. This test is based on a careful analysis of language performances that students are likely to engage in when studying at British universities. Careful descriptions of these tasks are provided as the input aspect of the test. Clearly, this approach is very different from the traditional approach where tests are constructed without this input foundation, usually by arbitrary decisions, and often just as results of correlations with other existing and usually bad tests. The input approach provides a careful analysis of language proficiency for the given purpose of testing. The tasks described in the TEAP are drawn from actual language performance, however, it is clear that these type of tasks by themselves cannot provide the necessary testing data until it has been proven that they also possess the necessary measurement criteria and requirements. Very little information is available to date on the output aspect of that test.

The paper presented by Baghdady in this meeting provides another example to prove that point. In his test there is also a very strong input component based on a systematic analysis and description of language proficiency, but there too what is lacking is the transformation of these tasks, most of which can qualify as good teaching tasks, into testing tasks.

The other type of tests that emerge are the output tests, which focus primarily on measurement properties. These tests are able to discriminate among various proficiency levels what they may be, they have high concurrent validity with other tests and high reliability. Tests such as the C-test which was described in this meeting, and in other IUS meetings by Klein-Braley and Raatz can be classified as output type of tests. The C-test is a very powerful instrument from a measurement point of view, it has excellent statistical properties - it discriminates among heterogeneous and homogeneous groups, between native and non-native speakers, among learners of different ages and levels and it has high inter-consistency reliability. And yet there

is no language learning theory to explain what it is doing and more importantly, why. The lack of language proficiency basis, the input, to serve as the foundation for the powerful output of the test, is definitely a limitation of that test. Supplying the seocnd half of every second word must be justified theoretically in terms of language learning. This is difficult to accept for these who share an expanded view of language testing.

In summary, these examples attempt to illustrate the two types of tests - on one hand tests which are strong on the input side and are based on an analysis of language proficiency but fail to prove their strength from a measurement point of view, and on the other hand, tests which are strong on the output side - pass various measurement criteria, but fail to provide a theory to explain these results, and therefore it is not really clear what they are measuring.

5. Conclusions

This paper criticized the approaches currently used for the development of language tests where some tests are based on a strong input aspect - what goes into a language test, and others are based on the output component - providing appropriate measurement criteria. It advocates the integration of both components, input and output, in order to produce good language tests. The paper recommends the descriptions provided by Stern and Cummins as a point of reference to focus on the input, while the output can be drawn from the current measurement theories. Needless to say, many of the tests which are available today do not possess either of these components.

It should be noted, however, that theories of language proficiency are still not complete and many of the sources of the input component have only basic preliminary information that could be used for test construction. At the same time measurement theories also still fail to provide sufficient and full direction regarding the psychometric procedures which should be used for dealing with language as it is defined today - communicative and authentic. In future years special efforts should be devoted to advancing these two components, input and output, in an integrated way into an "expanded" theory of language testing.

Bibliography

Blum-Kulka S. and E. Olshtain, 1984. Requests and Apologies: A Cross-Cultural Study of Speech-Act Realization Pattern, to appear in Applied Linguistics, Vol.5, No. 3

Stern H. and J. Cummins, Language Teaching/Language Research, 1980. In ACTFL Annual Review - Action for the 80s. National Textbook Company, Chicago, Illinois.

Stevenson Douglas K. Language, 1981. "Testing and Academic Accountability: On Redefining the Role of Language Testing in Language Teaching", IRAL, Vol. IX/1

Stevenson Douglas K., 1982. "Language Testing - All of the Above" in James C. (ed.) Research (working title) ACTFL Annual Review, Vol. 14, National Textbook Company, Skokie, Ill.